

Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review.

Year: 2021 | Citations: 282 | Authors: R. Daneshjou, Mary P Smith, Mary D Sun, V. Rotemberg, James Zou

Abstract

Importance

Clinical artificial intelligence (AI) algorithms have the potential to improve clinical care, but fair, generalizable algorithms depend on the clinical data on which they are trained and tested.

Objective

To assess whether data sets used for training diagnostic AI algorithms addressing skin disease are adequately described and to identify potential sources of bias in these data sets.

Data Sources

In this scoping review, PubMed was used to search for peer-reviewed research articles published between January 1, 2015, and November 1, 2020, with the following paired search terms: deep learning and dermatology, artificial intelligence and dermatology, deep learning and dermatologist, and artificial intelligence and dermatologist.

Study Selection

Studies that developed or tested an existing deep learning algorithm for triage, diagnosis, or monitoring using clinical or dermoscopic images of skin disease were selected, and the articles were independently reviewed by 2 investigators to verify that they met selection criteria.

Consensus Process

Data set audit criteria were determined by consensus of all authors after reviewing existing literature to highlight data set transparency and sources of bias.

Results

A total of 70 unique studies were included. Among these studies, 1 065 291 images were used to develop or test AI algorithms, of which only 257 372 (24.2%) were publicly available. Only 14 studies (20.0%) included descriptions of patient ethnicity or race in at least 1 data set used. Only 7 studies (10.0%) included any information about skin tone in at least 1 data set used. Thirty-six of the 56 studies developing new AI algorithms for cutaneous malignant neoplasms (64.3%) met the gold standard criteria for disease labeling. Public data sets were cited more often than private data sets, suggesting that public data sets contribute more to new development and benchmarks.

Conclusions and Relevance

This scoping review identified 3 issues in data sets that are used to develop and test clinical AI algorithms for skin disease that should be addressed before clinical translation: (1) sparsity of data set characterization and lack of transparency, (2) nonstandard and unverified disease labels, and (3) inability to fully assess patient diversity used for algorithm development and testing.